PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 3/00, C07H 15/12
C12N 15/00, A61K 39/12

A1

(11) International Publication Number: WO 93/23421

(43) International Publication Date: 25 November 1993 (25.11.93)

(21) International Application Number: PCT/US93/04365

(22) International Filing Date: 7 May 1993 (07.05.93)

(30) Priority data:
07/882,171 8 May 1992 (08.05.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 07/882,171 (CIP)
Filed on 8 May 1992 (08.05.92)

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; Corporate Patents - U.S., UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : MILLER, Timothy, J. [US/US]; 102 Crestside Way, Malvern, PA 19355 (US). KLEPFER, Sharon [US/US]; 113 Lindbergh Avenue, Broomall, PA 19008 (US). REED, Albert, Paul [US/US]; 117 Baker Circle, Exton, PA 19341 (US). JONES, Elaine, V. [US/US]; 1217 Andover Road, Wynnewood, PA 19096 (US).

(74) Agents: SCHRECK, Patricia, A. et al.; SmithKline Beecham Corporation, Corporate Patents - U.S., UW2220, 709 Swedeland Road, P.O. Box 1539, King of Pussia, PA 19406-0939 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: UNIVERSAL CORONAVIRUS VACCINE

(57) Abstract

A universal vaccine is disclosed which elicits a protective immune response in different host species and against different coronaviruses. A polypeptide which elicits protective antibodies against a homologous sequence found in the C terminal portion of coronavirus S proteins is disclosed. Vaccines comprising either the polypeptide or nucleic acids which encode the polypeptide are also disclosed. Methods of protecting a host against coronavirus infection are disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

Universal Coronavirus Vaccine

Cross reference to related applications

This application is a continuation-in-part application of U.S. application serial number 07/882,171, filed May 8, 1992, pending, which is a continuation-in-part of U.S. application serial number 07/698,927, filed May 13, 1991, which is a continuation-in-part of U.S. application serial number 07/613,066, filed November 14, 1990, each of which is incorporated herein by reference.

Field of the invention

The present invention relates to a universal vaccine useful to protect different species of animals against infection by different host-specific coronaviruses.

Background of the invention

Coronaviruses are a family of host-specific enveloped RNA viruses with a single-stranded positive sense genome. Examples of coronaviruses include, but are not limited to: feline infectious peritonitis (FIPV) and feline enteric coronavirus (FECV) which are specific to felines; canine coronavirus (CCV) which is specific to canines; transmissible gastroenteritis coronavirus (TGEV) which is specific to swine; bovine coronavirus (BCV) which is specific to bovine species; human coronavirus which is specific to humans; mouse hepatitis virus (MHV) which is specific to murine species; and infectious bronchitis virus (IBV) which is specific to avian species. These host-specific coronaviruses cannot cross infect different species of animals. Viral infection of the host by a coronavirus can cause symptoms ranging from mild enteritis to severe debilating disease to, in some cases, death.

Coronaviruses share common structural features including a spike or S protein (also referred to as a peplomer protein). The S protein is a glycoprotein which protrudes

from the surface of the virus particle. The S protein mediates the binding of virions to the host cell receptor and is involved in membrane fusion. In addition, it is the target of virus neutralizing antibodies.

S proteins contain an N-terminal signal sequence, a C-terminal transmembrane segment and potential N-linked glycosylation sites. Comparison of different coronavirus S proteins show little homology, i.e. similarity, at the N terminus and highly conserved amino acid sequences at the C terminus. Because the tissue tropism and disease symptomatology is quite varied among this virus family, it is speculated that the pathogenesis of coronaviruses is determined by the sequences encoded at the N-terminus while the more conserved C-terminus encodes critical structural features common to all coronaviruses. The carboxy terminus of the S protein is believed to be involved in fusion.

The structure of the S protein has been studied. Cavanagh (1983) *J. Gen. Virol.* **64**:2577-2583, which is incorporated herein by reference, proposed a model for the coronavirus spike in which the C-terminal half of the protein forms its stalk and the N-terminal half, its bulbous protein. deGroot *et al.*, (1987) *J. Mol. Biol.* **197**:, which is incorporated herein by reference, have postulated a model in which a coiled-coil structure forms the connection between the globular part of the S protein and the viral membrane. This model is based on the occurrence of heptad repeats, i.e., a periodicity (a-b-c-d-e-f-g) in which the amino acids are hydrophobic. Britton (1991) *Nature* **353**:394, which is incorporated herein by reference, reported the presence of a leucine zipper motif at the carboxyl end of the S glycoprotein of coronaviruses for which the spike sequence is available: TGEV FS772/70 (amino acids 1342-1377), FIPV WSU 1146 (amino acids 1345-1380), MHV A59 (amino acids 1217-1252), human coronavirus 229E (amino acids 1067-1102), BCV Mebus (amino acids 1266-1294), and infectious bronchitis virus Beaudette (amino acids 1059-1079). The leucine zipper motif terminates

ten residues upstream of the conserved KWP motif preceding the transmembrane domain.

Efforts have been made to develop vaccines against various host-specific coronaviruses. Attempts have been made with varying success to develop attenuated live virus vaccines, inactivated vaccines, subunit vaccines and recombinant nucleic acid based vaccines. In each case, the vaccine developed did not cross-protect other host animals. Vaccines currently available for protection against coronavirus are specific for protection against a given member of the coronavirus family. Such vaccines do not provide cross protection to protect a host against other members of the coronavirus family which are able to infect the species. Furthermore, such vaccines do not cross protect other animals against coronaviruses for which they are susceptible to infection.

There is a need for a vaccine which can protect against coronavirus infection. In particular, there is a need for a vaccine which can be useful to protect a host species against different coronaviruses and there is a need for a vaccine which can be useful to protect different host species against different coronaviruses.

Summary of the invention

The present invention relates to a polypeptide comprising an amino acid sequence from the C terminal portion of a coronavirus S protein which has been found to be highly conserved among coronaviruses and which is capable of eliciting a protective immune response. This sequence is referred to as a universal conserved domain. The polypeptides of the present invention have less than a complete amino acid sequence of an S protein.

The present invention relates to a vaccine comprising a polypeptide which includes an universal conserved domain and which has less than a complete amino acid sequence of an S protein.

The present invention relates to an isolated nucleic acid molecule having a nucleic acid sequence which encodes a polypeptide that includes a universal conserved domain polypeptide and that has less than a complete amino acid sequence of an S protein.

The present invention relates to a vaccine comprising a nucleic acid molecule that encodes a polypeptide which includes an universal conserved domain and which has less than a complete amino acid sequence of an S protein.

The present invention relates to a method of protecting an animal from infection by a coronavirus comprising administering an amount of a polypeptide effective to elicit a protective immune response. The polypeptide administered in the method comprises a universal conserved domain and has less than a complete amino acid sequence of an S protein.

The present invention relates to a method of protecting an animal from infection by a coronavirus comprising administering an amount of a nucleic acid molecule which encodes a polypeptide effective to elicit a protective immune response. The polypeptide encoded by the nucleic acid molecule administered in the method comprises a universal conserved domain and has less than a complete amino acid sequence of an S protein.

Detailed description of the invention

According to the present invention, a highly conserved region of the spike protein has been identified which, when presented as a vaccine component or product, is useful as a universal immunogen to protect an animal against coronavirus infection. The vaccine of the present invention may be used to vaccinate any animal susceptible to infection by virus that is a member of the coronavirus family. Accordingly, the present invention provides vaccines which can be produced in a single manufacturing process and administered to different species of animals. The cross-protection afforded by vaccines of the present invention eliminates the

need to produce different vaccines to protect animals against different members of the coronavirus family.

As used herein, the term "polypeptide" is meant to refer to a peptide, polypeptide or protein molecule; a molecule which includes a peptide, polypeptide or protein molecule; or a molecule that contains amino acid residues which are linked by non-peptide bonds.

As used herein, the term "universal conserved domain" ("UCD") is meant to refer to the identical 124 amino acid segment found in the C terminal portion of S proteins from TGEV, CCV and strains of feline coronaviruses. In addition, the term "UCD" is meant to refer to the corresponding amino acid segments of other coronavirus which have different but homologous amino acid sequences. Such corresponding sequences may be identified by their location in the S protein, i.e. downstream of the bulbous N-terminal region and upstream of the transmembrane region and the high level of amino acid sequence similarity to the 124 amino acid sequence described above. Furthermore, the term "UCD" is additionally meant to refer to consensus sequences are generated by comparing corresponding sequences and determining the statistically average amino acid residue at a given position in the sequence. Thus, when several different sequences are compared, the most common residue at a given position is assigned to that position in a consensus sequence.

The conservation of UCD sequences suggests that they play a major role in virus structure and/or replication. The region of perfect homology decreases in size as other coronavirus S genes are included in the comparison. For example, bovine and human coronavirus are more closely aligned to the feline, canine and porcine coronavirus S genes in this conserved region than are sequences from the murine and avian coronaviruses.

Table 1 contains a comparison of corresponding amino acid sequences from the C terminal portion of various coronaviruses. SEQ ID NO:1 is an amino acid sequence from FIPV strain Wsue2 (Virulent, Type II; Genbank accession number

X06170). SEQ ID NO:2 is an amino acid sequence from FIPV strain Df2e2 (Virulent, Type II). SEQ ID NO:3 is an amino acid sequence from FIPV strain Tse2 (Temperature sensitive mutant of Df2). SEQ ID NO:4 is an amino acid sequence from FECV strain Fecve2 (Avirulent strain 1683). SEQ ID NO:5 is an amino acid sequence from TGEV strain Tgeve2 (Purdue strain; Genbank accession number D00118). SEQ ID NO:6 is an amino acid sequence from FIPV strain Tgeve2f2 (Miller strain; Genbank accession number M56002). SEQ ID NO:7 is an amino acid sequence from BCV strain Bcve2 (Genbank accession number M30613). SEQ ID NO:8 is an amino acid sequence from HCV strain Hcve2 (Genbank accession number X16816). SEQ ID NO:9 is an amino acid sequence from IBV strain Ibbspi (Genbank accession number X16816). SEQ ID NO:10 is an amino acid sequence from MHV strain Mhve2a59 (Genbank accession number X51939 SEQ ID NO:11 is an amino acid sequence from FIPV strain Mhvs (Genbank accession number X04797). SEQ ID NO:12 is a consensus sequence which has been designed to provide an optimum UCD amino acid sequence.

The 124 residue amino acid sequence which is completely conserved in TGEV, CCV and feline coronaviruses is shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:5 from residue 37 to residue 160. The consensus sequence, SEQ ID NO:12, also contains this 124 amino acid sequence in its entirety from residue 37 to residue 160. This 124 amino acid sequence is currently a preferred UCD sequence of the present invention. The entire 199 amino acid consensus sequence is a preferred UCD-containing peptide.

Using amino acid sequence information from any coronavirus, one having ordinary skill in the art can identify the conserved region corresponding to the 124 amino acid sequence found in TGEV, CCV and feline coronaviruses. As exemplified in Table 1, the amino acid sequences from the C terminal portion of coronaviruses can be compared to identify the sequence which corresponds to the UCD from TGEV, CCV and feline coronaviruses. The procedure is straightforward and

can be performed to provide additional UCD sequences and flanking sequences.

Corresponding conserved regions from coronaviruses other than CCV, TGEV and feline coronaviruses may be identified by their location on the S protein and the high level of sequence homology the possess when compared to the 124 amino acid sequence referred to above. An example of such comparison and identification is shown in Table 1 in which sequences from the C terminal regions of various S proteins upstream from the transmembrane region are compared and homologous sequences identified. Widely available computer programs such as PLOTSIMILARITY software (Genetics Computer Group, Madison WI) may be employed to locate a UCD in a coronavirus.

In addition, such software may be employed to expedite the generation of consensus sequences. This software relies on the principles originally set out by Wilbur and Lipman and later refined by Smith and Waterman and by Needleman and Wunsch. Using these well known guidelines, having ordinary skill in the art may compare sequences and arrive at the statistically average or most common residue occupying a given position. The PLOTSIMILARITY software automates this function. Consensus sequences are thus generated. In addition to the consensus sequence provided as SEQ ID NO:12, a different consensus sequence derived from a comparison of corresponding sequences is disclosed in the co-owned, co-pending patent application: which is filed on the same day as the present application; which is entitled "Compositions and Methods for Vaccinating Coronaviruses"; which names the same inventors as the present application (Miller, Timothy J.; Jones, Elaine V.; Reed, Albert P.; and Klepfer, Sharon R); which has been designated docket number H85009-1 by Applicants; and which is incorporated herein by reference.

Accordingly, the present invention relates to polypeptides which comprise a UCD or a fragment or a derivative thereof. That is, the present invention relates

to polypeptides which comprise: the 124 amino acid sequence form TGEV, CCV and feline coronaviruses; or the different amino acid sequences from other coronaviruses which correspond to the 124 amino acid sequence; or a consensus sequence generated from comparison of corresponding regions; or immunogenic fragments or immunogenic derivatives thereof.

Polypeptides according to the present may further comprise additional flanking sequences from coronavirus or flanking sequences designed as a consensus sequence of the flanking sequences of corresponding regions from different coronaviruses.

As used herein, the term "immunogenic fragment" is meant to refer to polypeptides which include an incomplete UCD which is capable of eliciting a protective immune response against coronavirus in an animal susceptible to coronavirus infection. Immunogenic fragments may comprise a sequence having nine or more amino acids from a UCD, and may include additional amino acid sequences.

As used herein, the term "immunogenic derivatives" is meant to refer to molecules which have a UCD or portions thereof with conservative amino acid substitutions and which are capable of eliciting a protective immune response against a coronavirus in an animal susceptible to coronavirus infection. Those having ordinary skill in the art can readily design derivatives having UCD sequences with conservative substitutions for amino acids. For example, following what are referred to as Dayhof's rules for amino acid substitution (Dayhof, M.D. (1978) *Nat. Biomed. Res. Found.*, Washington, D.C. Vol. 5, supp. 3), amino acid residues in a peptide sequence may be substituted with comparable amino acid residues. Such substitutions are well known and are based the upon charge and structural characteristics of each amino acid.

Using standard procedures and readily available starting materials, one having ordinary skill in the art can determine whether a fragment and derivative is an immunogenic fragment or an immunogenic derivative, respectively. Briefly, polypeptides can be produced by standard methodologies and

tested to determine whether they are capable of eliciting a protective immune response. Sera from vaccinated animals can be analyzed to detect the presence of antibodies capable of inhibiting infection of cells in culture. Furthermore, challenge studies can be performed to determine if animals vaccinated with a polypeptide are protected from subsequent infection by wild type virus. One having ordinary skill in the art can routinely produce and screen fragments and derivatives to determine the effectiveness of such vaccine components to elicit protective immune responses. Similarly, larger molecules may also be screened by the same means to detect their ability to elicit a protective immune response.

The UCD lies near the transmembrane region of the S protein. Because this region of the S protein is purported to be involved in the secondary structure of the glycoprotein, in receptor binding and in virus-induced cell fusion, the UCD plays an important role in the function of the S protein and in the formation of infectious virus. Inducing an immune response against this region will interfere with the folding of the S glycoprotein into its proper conformation. The presence of circulating antibodies to this region could bind to either virus or infected cells expressing the glycoprotein on the surface. Virus complexed with antibody may be unable to bind to receptors on susceptible cells and/or initiate the pathway required to gain entry which involves a conformational change of the S protein. Recognition of this region on the surface of infected cells would target them for clearance. Antibody binding to the conserved region of the S protein surface expressed by infected cells would, most likely, prevent cell fusion and interfere with virus assembly. Regardless of mechanism, an immune response to the UCD of a coronavirus S protein will inhibit virus spread from cell to cell and limit virus infection.

Polypeptides according to the present invention comprise less than a complete S protein sequence. In particular, the polypeptides do not comprise a complete N-terminal portion of an S protein and preferably comprise few

or no amino acid sequences from the N-terminal bulbous portion of the protein. Furthermore, the polypeptides preferably do not comprise a complete transmembrane domain of an S protein. In some preferred embodiments, polypeptides comprise no more than a 400 amino acid sequence upstream (from the C terminus to the N terminus) from about 2 amino acids upstream from the transmembrane domain. In some preferred embodiments, polypeptides comprise no more than a 300 amino acid sequence upstream (from the C terminus to the N terminus) from about 5 amino acids upstream from the transmembrane domain.

In some preferred embodiments, polypeptides which comprise a UCD, or derivatives and/or fragments thereof further comprise flanking sequences of the UCD found in coronavirus. For example, in some preferred embodiments, the polypeptide comprises portions of the S protein flanked by and optionally including the heptad repeats reported by deGroot *et al.*, such as, for example, in FIPV strain WSU 1146 from residues 1067 to 1380. In some preferred embodiments, the polypeptide comprises portions of the S protein flanked on the carboxy side by and may also include a leucine zipper motif as reported by Britton. In some preferred embodiments, the polypeptide comprises portions of the S protein from about 300 residues upstream of the transmembrane region to about 5 amino acid residues upstream from the transmembrane domain.

In some preferred embodiments, the polypeptide comprises a UCD about 124 amino acids in length. In some preferred embodiments, the polypeptide comprises an immunogenic fragment of a UCD about 100 amino acids in length. In some preferred embodiments, the polypeptide comprises an immunogenic fragment of a UCD about 50 amino acids in length. In some preferred embodiments, the polypeptide comprises an immunogenic fragment of a UCD about 25 amino acids in length. In some preferred embodiments, the polypeptide comprises an immunogenic fragment of a UCD about 15 amino acids in length. In some preferred embodiments, the polypeptide comprises an immunogenic fragment of a UCD about 10 amino acids in length.

In some preferred embodiments, a UCD comprises amino acid residues 37-160 of SEQ ID NO:12. Additional preferred embodiments comprise SEQ ID NO:12. Other preferred embodiments of the invention comprise SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 or SEQ ID NO:5. Other preferred embodiments comprise SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 or SEQ ID NO:11.

In addition to a UCD and, optionally, additional flanking segments from an S protein, other peptide segments may also be included in the polypeptide of the present invention. Such additional peptide segments may comprise other immunogenic targets from coronavirus and/or other pathogens, and/or they may be provided for improved stability, UCD epitope presentation or production/purification facilitation. The resulting polypeptide is considered a chimeric or fusion polypeptides.

Vaccines according to the present invention can be employed to vaccinate animals against infection by coronaviruses or at least to prevent the clinical symptoms associated with such infections. Such vaccines will provide protection against multiple coronaviruses and cross species protection. Vaccines may be produced which are either protein-based or nucleic acid-based. In both cases, the vaccinated animal is exposed to an immunogenic polypeptide which comprises a UCD. A protective immune response is elicited which is sufficient to protect the animal against coronavirus.

Vaccines according to the present invention can be either:

a) compositions which comprise a polypeptide that includes a universal conserved domain; or

b) compositions which comprise a nucleic acid molecule that includes a nucleotide sequence which encodes a polypeptide that includes a universal conserved domain. In both types of vaccines, the polypeptide is not a complete S protein and it elicits a protective immune response in animals.

In protein based, i.e. subunit vaccines, polypeptides having a UCD may by produced using standard techniques including recombinant DNA techniques for protein production or by peptide synthesis. In preferred embodiments, polypeptides used in subunit vaccines according to the present invention are produced by recombinant DNA methodology.

The nucleic acid sequences of coronavirus S genes are widely known. One having ordinary skill in the art may routinely obtain DNA that encodes a polypeptide including a UCD using standard techniques and widely available starting materials. The nucleotide and amino acid sequences for S proteins from several types and strains of coronaviruses can be found in the co-owned published PCT application PCT/US91/08525 which claims priority to U.S. Patent Application Serial Numbers 613,066 and 698,927; each of these applications are incorporated herein by reference. Nucleotide and amino acid sequences of S proteins can also be found in published European Patent Applications publication numbers: 0,524,672 A1; 0,411,684 A2; 0,264,979 A1; 0,138,242 A1; and application number EP 91 30 3737. Each of these European patent applications are incorporated herein by reference. In addition, nucleotide and amino acid sequences of S proteins from several coronaviruses as well as nucleotide and amino acid sequences of a consensus sequence is disclosed in the co-owned, co-pending patent application: which is filed on the same day as the present application; which is entitled "Compositions and Methods for Vaccinating Coronaviruses"; which names the same inventors as the present application (Miller, Timothy J.; Jones, Elaine V.; Reed, Albert P.; and Klepfer, Sharon R); which has been designated docket number H85009-1 by Applicants; and which is incorporated herein by reference.

Nucleic acid molecules encoding some or all of an S protein from a coronavirus may be generated by a variety of techniques. For such molecules, a nucleotide sequence that encodes a UCD may be identified. Using, for example, Polymerase Chain Reaction (PCR) methodology, primers flanking

both sides the region of interest may be designed and used to produce multiple copies of the UCD routinely. Alternatively, using restriction enzymes, a UCD may be isolated from DNA encoding an S protein. Moreover, nucleic acid molecules that encode a UCD may also be synthesized using techniques well known to those having ordinary skill in the art.

One having ordinary skill in the art can, using well known techniques, insert such DNA molecules into a commercially available expression vector for use in well known expression systems. For example, the commercially available plasmid pSE420 (Invitrogen, San Diego, CA) may be used for production of a DNA encoding a polypeptide including a UCD in *E. coli*. The commercially available plasmid pYES2 (Invitrogen, San Diego, CA) may, for example, be used for production in *S. cerevisiae* strains of yeast. The commercially available MaxBac™ (Invitrogen, San Diego, CA) complete baculovirus expression system may, for example, be used for production in insect cells. The commercially available plasmid pcDNA I (Invitrogen, San Diego, CA) may, for example, be used for production in mammalian cells such as Chinese Hamster Ovary cells. One having ordinary skill in the art can use these commercial expression vectors and systems or others to produce a polypeptide including a UCD using routine techniques and readily available starting materials. (*See e.g.*, Sambrook et al., *Molecular Cloning a Laboratory Manual*, Second Ed. Cold Spring Harbor Press (1989) which is incorporated herein by reference.) Thus, the desired proteins can be prepared in both prokaryotic and eukaryotic systems, resulting in a spectrum of processed forms of the protein.

The particulars for the construction of expression systems suitable for desired hosts are known to those in the art. Briefly, for recombinant production of the protein, the DNA encoding the polypeptide is suitably ligated into the expression vector of choice. The DNA is operably linked to all regulatory elements which are necessary for expression of the DNA in the selected host. One having ordinary skill in

the art can, using well known techniques, prepare expression vectors for recombinant production of the polypeptide.

The expression vector including the DNA that encodes the polypeptide comprising a UCD is used to transform the compatible host which is then cultured and maintained under conditions wherein expression of the foreign DNA takes place. The protein of the present invention thus produced is recovered from the culture, either by lysing the cells or from the culture medium as appropriate and known to those in the art. One having ordinary skill in the art can, using well known techniques, isolate the polypeptide that includes a UCD produced using such expression systems.

In addition to producing these proteins by recombinant techniques, automated peptide synthesizers may also be employed to produce polypeptides that include a UCD. Such techniques are well known to those having ordinary skill in the art and are useful if derivatives which have substitutions not provided for in DNA-encoded protein production.

Subunit vaccines according to the invention comprise a polypeptide the includes a UCD but which is not a complete S protein and a pharmaceutically acceptable carrier or diluent. Optionally, the vaccine may comprise additional immunogenic proteins, additional vaccine components such as non-subunit vaccines, and/or an adjuvant.

In nucleic acid molecule based, i.e. recombinant vaccines, a nucleotide sequences which encode polypeptides that include a UCD is inserted into a vector and administered to the animal. The vector delivers genetic material to the animal where it is transcribed and translated to produce the immunogenic polypeptide. Vectors for use as vaccines are well known and include non-pathogenic viruses and prokaryotic organisms. Suitable vectors for delivering genetic material are readily available or may be produced from readily available starting materials using standard techniques. Two examples of vectors useful for delivering genetic material as a vaccine are the recombinant pox vectors or non-pathogenic

Salmonella strains. The nucleotide sequence that encodes the immunogenic polypeptide is operably linked to regulatory elements required for expression and inserted within the vector. Alternatively, it is incorporated into the vector at a site where it is placed under the control of the necessary regulatory elements already present in the vector. Naked DNA may also be used as a vaccine delivery system.

Recombinant vaccines may be used in combination with other vaccines. Further, the genetic material which encodes the polypeptide that comprises the UCD may further comprise additional coding sequences which encode other peptide sequences capable of eliciting an immunogenic response against coronavirus or another pathogen.

Both subunit and recombinant vaccines may be formulated following accepted convention using buffers, stabilizers, preservative, solubilizers and compositions used to facilitate sustained release. Generally, additives for isotonicity can include sodium chloride, dextrose, mannitol, sorbitol and lactose. Stabilizers include gelatin and albumin. Adjuvants such as aluminum or magnesium hydroxide may be employed. Vaccines may be maintained in solution or, in some cases, particularly recombinant vaccines, lyophilized. Lyophilized vaccine may be stored conveniently and combined with sterile solution before administration.

The amount of polypeptide administered depends upon such factors as the size of the polypeptide, the species, age, weight, and general physical characteristics of the animal, and by the composition of the vaccine. Determination of optimum dosage for each parameter may be made by routine methods. Generally, subunit vaccines according to the present invention contain between 0.05-5000 micrograms of polypeptide per milliliter of sterile solution, preferably 10-1000 micrograms. Generally, recombinant vaccines according to the present invention contain between 10^5-10^8 infectious units per milliliter of sterile solution. About .5-2 milliliter of polypeptide-containing solution is administered.

Subunit vaccines and genetic material based vaccines may be administered by an appropriate route such as, for example, by oral, intranasal, intramuscular, intraperitoneal or subcutaneous administration. In some embodiments, intranasal or subcutaneous administration is preferred. Subsequent to initial vaccination, animals may be boosted by revaccination.

Examples

Example 1 Cloning of Coronavirus Conserved Region in pMG1

The bacterial expression vector, pMG-1, allows a gene expressing a foreign protein to be fused to a partial sequence of the NS1 gene from influenza virus, the first 81 encoding amino acids thereof. This vector is described in European Patent Application No. 366,238, published May 2, 1990, which is incorporated herein by reference.

Primers were designed to amplify a S gene region encoding amino acids 1115-1238 of the DF2 FIPV strain for expression in this vector as follows. The upstream primer contains NcoI and NdeI restriction sites and initiates amplification at base pair 3406 (amino acid 1115), and is SEQ ID NO:13:

5'-
GTTGTCAACACACCATGGATCATATGCAAGGGCAAGCTTTAAGTCACCTTACA.
NcoI NdeI

The downstream primer contains a StuI site and terminates amplification at base pair 3777 (amino acid 1238), and is SEQ ID NO: 14:

5'-AAATACCTGAGGCCTCCAAGCTGTTACAGTTTCATAAGCTGT.
StuI

The amplified fragment (412 bp) was cloned into the pT_7 Blue vector according to the manufacturer's instructions. A plasmid containing amino acids 1115-1238 in pT_7 Blue was digested with NcoI/StuI, the 412 base pair insert isolated, and ligated overnight at 15°C to plasmid vector pMG1 digested with NcoI/StuI and dephosphorylated. Host cells AR120 and AR58 were transformed with the ligation mix and the presence

of insert bearing clones was confirmed by diagnostic restriction enzyme digestions.

Example 2 - Cloning of Coronavirus Conserved Region in pSC11

Vaccinia recombinants were engineered to contain the 1115-1238 amino acid conserved region of WT DF2 FIPV. The conserved region was cloned into the vaccinia expression vector pSC11 by blunt-ending the 412 base pairs NcoI/StuI fragment isolated from the pT7 Blue clone described in Example 12, end-filling by incubation with Klenow polymerase, and inserting it into the SmaI site downstream of the 7.5K vaccinia promoter. The ligation mix was transformed into HB101 host cells. Full-length clones were identified and oriented with respect to vector by BamHI and ScaI digests of mini-prep DNAs, respectively.

Table 1

```
                1                                                   50
Wsue2           NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Df2e2           NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Tse2            NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Fecve2          NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Tgeve2          NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Tgeve2f2        NITQAFGKVN DAIHQTSQGL ATVAKALAKV QDVVNTQGQA LSHLTVQLQN
Bcve2           AIQEGFDATN S.......... .....ALVKI QAVVNANAEA LNNLLQQLSN
Hcve2           NIVDAFTGVN DAITQTSQAL QTVATALNKI QDVVNQQGNS LNHLTSQLRQ
Ibbspi          HMQE...... ........GF RSTSLALQQI QDVVSKQSAI LTETMASLNK
Mhve2a59        AIQDGFDATN S.......... .....ALGKI QSVVNANAEA LNNLLNQLSN
Mhvs            AIQEGFDATN S.......... .....ALGKI QSVVNANAEA LNNLLNQLSN
CONSENSUS       NITQAFGKVN DAIHQTS.GL ATVAKALAKV QDVVNTQGQA LSHLTVQLGN

                51                                                  100
Wsue2           NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Df2e2           NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Tse2            NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Fecve2          NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Tgeve2          NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Tgeve2f2        NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV
Bcve2           RFGAISSSLQ EILSRLDALE AQAQIDRLIN GRLTALNVYV SQQLSDSTLV
Hcve2           NFQAISSSIQ AIYDRLDTIQ ADQQVDRLIT GRLAALNVFV SHTLTKYTEV
Ibbspi          NFGAISSVIQ EIUQQFDAIQ ANAQVDRLIT GRLSSLSVLA SAKQAEUIRV
Mhve2a59        RFGAISASLQ EILTRLEAVE AKAQIDRLIN GRLTALNAYI SKQLSDSTLI
Mhvs            RFGAISASLQ EILTRLDAVE AKAQIDRLIN GRLTALNAYI SKQLSDSTLI
CONSENSUS       NFQAISSSIS DIYNRLDELS ADAQVDRLIT GRLTALNAFV SQTLTRQAEV

                101                                                 150
Wsue2           RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Df2e2           RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Tse2            RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Fecve2          RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Tgeve2          RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Tgeve2f2        RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL
Bcve2           KFSAAQAMEK VNECVKSQSS RINFCGNGNH IISLVQNAPY GLYFIHFSYV
Hcve2           RASRQLAQQK VNECVKSQSK RYGFCGNGTH IFSIVNAAPE GLVFLHTVLL
Ibbspi          SQQRELATQK INECVKSQSI RYSFCGNGRH VLTIPQNAPN GIVFIHFSYT
Mhve2a59        KVSAAQAIEK VNECVKSQTT RINFCGNGNH ILSLVQNAPY GLYFIHFSYV
Mhvs            KFSAAQAIEK VNECVKSQTT RINFCGNGNH ILSLVQNAPY GLCFIHFSYV
CONSENSUS       RASRQLAKDK VNECVRSQSQ RFGFCGNGTH LFSLANAAPN GMIFFHTVLL

                151                                                 200
Wsue2           PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Df2e2           PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Tse2            PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Fecve2          PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Tgeve2          PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Tgeve2f2        PTAYETVTAW SGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
Bcve2           PTKYVTAKYS PGLCIA.GDR GIA.....PK SGYFVNVNNT WMFTGSGYYY
Hcve2           PTQYKDVEAW SGLC..:VDG TNGYVLRQPN LALYKE.GNY YRITSRIMFE
Ibbspi          PDSFVNVTAI VGFCVKPANA SQUAIVPANG RGIFIQVNGS YYITARDMYM
Mhve2a59        PISFTTANVS PGLCIS.GDR GLA.....PK AGYFVQDDGE WKFTGSSYYY
Mhvs            PTSFKTANVS PGLCIS.GDR GLA.....PK AGYFVQDNGE WKFTGSNYYY
CONSENSUS       PTAYETVTAW PGICASDGDR TFGLVVKDVQ LTLFRNLDDK FYLTPRTMYQ
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Miller, Timothy J.
Jones, Elaine V.
Reed, Albert P.
Klepfer, Sharon R.

(ii) TITLE OF INVENTION: Universal Coronavirus Vaccine

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19406-2799

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/882,171
(B) FILING DATE: 08-MAY-1992

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/698,927
(B) FILING DATE: 13-MAY-1991

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/613,066
(B) FILING DATE: 14-NOV-1990

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Schreck, Patricia A.
(B) REGISTRATION NUMBER: 33,777
(C) REFERENCE/DOCKET NUMBER: SBC/PAS/WW001

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn.
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 200 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gln Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp
20 25 30

Val Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu
35 40 45

Gln Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn
50 55 60

Arg Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg
85 90 95

Gln Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn
100 105 110

Glu Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe
130 135 140

Phe His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp
145 150 155 160

Ser Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val
165 170 175

Lys Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr
180 185 190

Leu Thr Pro Arg Thr Met Tyr Gln
195 200

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 179 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Ala Ile Gln Glu Gly Phe Asp Ala Thr Asn Ser Ala Leu Val Lys Ile
1 5 10 15

Gln Ala Val Val Asn Ala Asn Ala Glu Ala Leu Asn Asn Leu Leu Gln
20 25 30

Gln Leu Ser Asn Arg Phe Gly Ala Ile Ser Ser Ser Leu Gln Glu Ile
35 40 45

Leu Ser Arg Leu Asp Ala Leu Glu Ala Gln Ala Gln Ile Asp Arg Leu
50 55 60

Ile Asn Gly Arg Leu Thr Ala Leu Asn Val Tyr Val Ser Gln Gln Leu
65 70 75 80

Ser Asp Ser Thr Leu Val Lys Phe Ser Ala Ala Gln Ala Met Glu Lys
85 90 95

Val Asn Glu Cys Val Lys Ser Gln Ser Ser Arg Ile Asn Phe Gly Asn
100 105 110

Gly Asn His Ile Ile Ser Leu Val Gln Asn Ala Pro Tyr Gly Leu Tyr
115 120 125

Phe Ile His Phe Ser Tyr Val Pro Thr Lys Tyr Val Thr Ala Lys Tyr
130 135 140

Ser Pro Gly Leu Cys Ile Ala Gly Asp Arg Gly Ile Ala Pro Lys Ser
145 150 155 160

Gly Tyr Phe Val Asn Val Asn Asn Thr Trp Met Phe Thr Gly Ser Gly
165 170 175

Tyr Tyr Tyr

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 196 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asn Ile Val Asp Ala Phe Thr Gly Val Asn Asp Ala Ile Thr Gln Thr
1 5 10 15

Ser Gln Ala Leu Gln Thr Val Ala Thr Ala Leu Asn Lys Ile Gln Asp
20 25 30

Val Val Asn Gln Gln Gly Asn Ser Leu Asn His Leu Thr Ser Gln Leu
35 40 45

Arg Gln Asn Phe Gln Ala Ile Ser Ser Ser Ile Gln Ala Ile Tyr Asp
50 55 60

Arg Leu Asp Thr Ile Gln Ala Asp Gln Gln Val Asp Arg Leu Ile Thr
65 70 75 80

Gly Arg Leu Ala Ala Leu Asn Val Phe Val Ser His Thr Leu Thr Lys
85 90 95

Tyr Thr Glu Val Arg Ala Ser Arg Gln Leu Ala Gln Gln Lys Val Asn
100 105 110

Glu Cys Val Lys Ser Gln Ser Lys Arg Tyr Gly Phe Cys Gly Asn Gly
115 120 125

Thr His Ile Phe Ser Ile Val Asn Ala Ala Pro Glu Gly Leu Val Phe
130 135 140

Leu His Thr Val Leu Leu Pro Thr Gln Tyr Lys Asp Val Glu Ala Trp
145 150 155 160

Ser Gly Leu Cys Val Asp Gly Thr Asn Gly Tyr Val Leu Arg Gln Pro
165 170 175

Asn Leu Ala Leu Tyr Lys Glu Gly Asn Tyr Tyr Arg Ile Thr Ser Arg
180 185 190

Ile Met Phe Glu
195

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
His Met Gln Glu Gly Phe Arg Ser Thr Ser Leu Ala Leu Gln Gln Ile
1               5                  10                  15

Gln Asp Val Val Ser Lys Gln Ser Ala Ile Leu Thr Glu Thr Met Ala
            20                  25                  30

Ser Leu Asn Lys Asn Phe Gly Ala Ile Ser Ser Val Ile Gln Glu Ile
        35                  40                  45

Gln Gln Phe Asp Ala Ile Gln Ala Asn Ala Gln Val Asp Arg Leu Ile
    50                  55                  60

Thr Gly Arg Leu Ser Ser Leu Ser Val Leu Ala Ser Ala Lys Gln Ala
65                  70                  75                  80

Glu Ile Arg Val Ser Gln Gln Arg Glu Leu Ala Thr Gln Lys Ile Asn
                85                  90                  95

Glu Cys Val Lys Ser Gln Ser Ile Arg Tyr Ser Phe Cys Gly Asn Gly
            100                 105                 110

Arg His Val Leu Thr Ile Pro Gln Asn Ala Pro Asn Gly Ile Val Phe
        115                 120                 125

Ile His Phe Ser Tyr Thr Pro Asp Ser Phe Val Asn Val Thr Ala Ile
    130                 135                 140

Val Gly Phe Cys Val Lys Pro Ala Asn Ala Ser Gln Ala Ile Val Pro
145                 150                 155                 160

Ala Asn Gly Arg Gly Ile Phe Ile Gln Val Asn Gly Ser Tyr Tyr Ile
                165                 170                 175

Thr Ala Arg Asp Met Tyr Met
            180
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 180 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ala Ile Gln Asp Gly Phe Asp Ala Thr Asn Ser Ala Leu Gly Lys Ile
1               5                  10                  15

Gln Ser Val Val Asn Ala Asn Ala Glu Ala Leu Asn Asn Leu Leu Asn
            20                  25                  30

Gln Leu Ser Asn Arg Phe Gly Ala Ile Ser Ala Ser Leu Gln Glu Ile
        35                  40                  45

Leu Thr Arg Leu Glu Ala Val Glu Ala Lys Ala Gln Ile Asp Arg Leu
    50                  55                  60
```

Ile Asn Gly Arg Leu Thr Ala Leu Asn Ala Tyr Ile Ser Lys Gln Leu
65 70 75 80

Ser Asp Ser Thr Leu Ile Lys Val Ser Ala Ala Gln Ala Ile Glu Lys
85 90 95

Val Asn Glu Cys Val Lys Ser Gln Thr Thr Arg Ile Asn Phe Cys Gly
100 105 110

Asn Gly Asn His Ile Leu Ser Leu Val Gln Asn Ala Pro Tyr Gly Leu
115 120 125

Tyr Phe Ile His Phe Ser Tyr Val Pro Ile Ser Phe Thr Thr Ala Asn
130 135 140

Val Ser Pro Gly Leu Cys Ile Ser Gly Asp Arg Gly Leu Ala Pro Lys
145 150 155 160

Ala Gly Tyr Phe Val Gln Asp Asp Gly Glu Trp Lys Phe Thr Gly Ser
165 170 175

Ser Tyr Tyr Tyr
180

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 180 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Ala Ile Gln Glu Gly Phe Asp Ala Thr Asn Ser Ala Leu Gly Lys Ile
1 5 10 15

Gln Ser Val Val Asn Ala Asn Ala Glu Ala Leu Asn Asn Leu Leu Asn
20 25 30

Gln Leu Ser Asn Arg Phe Gly Ala Ile Ser Ala Ser Leu Gln Glu Ile
35 40 45

Leu Thr Arg Leu Asp Ala Val Glu Ala Lys Ala Gln Ile Asp Arg Leu
50 55 60

Ile Asn Gly Arg Leu Thr Ala Leu Asn Ala Tyr Ile Ser Lys Gln Leu
65 70 75 80

Ser Asp Ser Thr Leu Ile Lys Phe Ser Ala Ala Gln Ala Ile Glu Lys
85 90 95

Val Asn Glu Cys Val Lys Ser Gln Thr Thr Arg Ile Asn Phe Cys Gly
100 105 110

Asn Gly Asn His Ile Leu Ser Leu Val Gln Asn Ala Pro Tyr Gly Leu
115 120 125

Cys Phe Ile His Phe Ser Tyr Val Pro Thr Ser Phe Lys Thr Ala Asn
130 135 140

Val Ser Pro Gly Leu Cys Ile Ser Gly Asp Arg Gly Leu Ala Pro Lys
145 150 155 160

Ala Gly Tyr Phe Val Gln Asp Asn Gly Glu Trp Lys Phe Thr Gly Ser
165 170 175

Asn Tyr Tyr Tyr
180

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asn Ile Thr Gln Ala Phe Gly Lys Val Asn Asp Ala Ile His Gln Thr
1 5 10 15

Ser Gly Leu Ala Thr Val Ala Lys Ala Leu Ala Lys Val Gln Asp Val
20 25 30

Val Asn Thr Gln Gly Gln Ala Leu Ser His Leu Thr Val Gln Leu Gly
35 40 45

Asn Asn Phe Gln Ala Ile Ser Ser Ser Ile Ser Asp Ile Tyr Asn Arg
50 55 60

Leu Asp Glu Leu Ser Ala Asp Ala Gln Val Asp Arg Leu Ile Thr Gly
65 70 75 80

Arg Leu Thr Ala Leu Asn Ala Phe Val Ser Gln Thr Leu Thr Arg Gln
85 90 95

Ala Glu Val Arg Ala Ser Arg Gln Leu Ala Lys Asp Lys Val Asn Glu
100 105 110

Cys Val Arg Ser Gln Ser Gln Arg Phe Gly Phe Cys Gly Asn Gly Thr
115 120 125

His Leu Phe Ser Leu Ala Asn Ala Ala Pro Asn Gly Met Ile Phe Phe
130 135 140

His Thr Val Leu Leu Pro Thr Ala Tyr Glu Thr Val Thr Ala Trp Pro
145 150 155 160

Gly Ile Cys Ala Ser Asp Gly Asp Arg Thr Phe Gly Leu Val Val Lys
165 170 175

Asp Val Gln Leu Thr Leu Phe Arg Asn Leu Asp Asp Lys Phe Tyr Leu
180 185 190

Thr Pro Arg Thr Met Tyr Gln
195

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTTGTCAACA CACCATGGAT CATATGCAAG GGCAAGCTTT AAGTCACCTT ACA 53

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAATACCTGA GGCCTCCAAG CTGTTACAGT TTCATAAGCT GT 42

Claims

1. A polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide having less than a complete amino acid sequence of said S protein.

2. A vaccine comprising a pharmaceutically acceptable carrier or diluent and a polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide having less than a complete amino acid sequence of said S protein.

3. A nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide having less than a complete amino acid sequence of said S protein.

4. A recombinant vaccine comprising a nucleic acid molecule, said nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide having less than a complete amino acid sequence of said S protein.

5. A method of protecting an animal against coronavirus comprising administering a polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide having less than a complete amino acid sequence of said S protein.

6. A method of protecting an animal against coronavirus comprising administering a nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide comprising a universal conserved domain of a coronavirus or an immunogenic fragment or derivative thereof; said polypeptide

having less than a complete amino acid sequence of said S protein.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/04365

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07K 3/00; C07H 15/12; C12N 15/00; A61K 39/12
US CL :530/350; 536/27; 435/320.1; 424/89
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 536/27; 435/320.1; 424/89

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP, A, 0,264,979 (deGroot et al) 27 April 1988, see entire document.	1-6
Y	Virus Research, Volume 8, issued 1987, L. Jacobs et al., "The Nucleotide Sequence of the Peplomer Gene of Porcine Transmissable Gastroenteritis Virus (TGEV): Comparison with the Sequence of the Peplomer Protein of Feline Infectious Peritonitis Virus (FIPV)", pp. 363-371, see entire document.	1-6

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 12 JULY 1993	Date of mailing of the international search report 03 AUG 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer D. BARND Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/04365

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	The Journal of General Virology, Volume 71, No. 5, issued May 1990, T. Raabe et al., "Nucleotide Sequence of the Gene Encoding the Spike Glycoprotein of Human Coronavirus HCV 229E", pp. 1065-1073, see entire document.	1-6
Y	Archives of Virology, Volume 117, issued 1991, T. Hohdatsu et al., "Characterization of Monoclonal Antibodies Against Feline Infectious Peritonitis Virus Type II and Antigenic Relationship Between Feline, Porcine, and Canine Coronaviruses", pp. 85-95, see entire document.	1-6
Y	Virology, Volume 174, No. 2, issued February 1990, C. Sanchez et al., "Antigenic Homology Among Coronaviruses Related to Transmissable Gastroenteritis Virus", pp. 410-417, see entire document.	1-6

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

EMBL, GenBank, GeneSeq, PIR, Swiss-Prot, CA, Biosis, Medline, Embase, WPI, APS
search terms: coronavirus, conserv?, spike, peplomer, C-term?, vaccine